

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-091484

(43)Date of publication of application : 27.03.2002

(51)Int.Cl.

G10L 15/18

(21)Application number : 2000-280655

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 14.09.2000

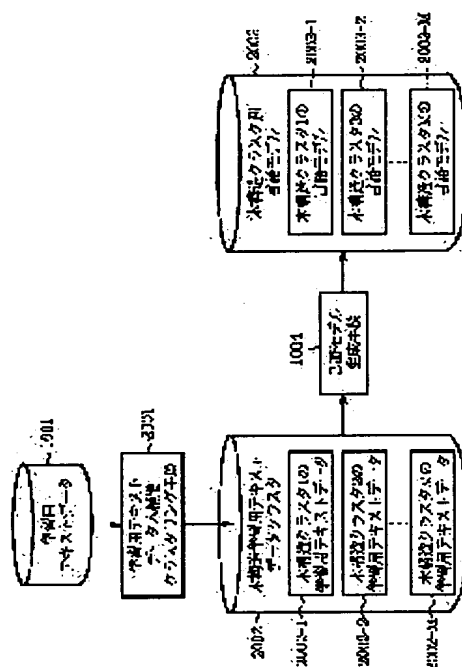
(72)Inventor : ISHII JUN

(54) LANGUAGE MODEL GENERATOR AND VOICE RECOGNITION DEVICE USING THE GENERATOR, LANGUAGE MODEL GENERATING METHOD AND VOICE RECOGNITION METHOD USING THE METHOD, COMPUTER READABLE RECORDING MEDIUM WHICH RECORDS LANGUAGE MODEL GENERATING PROGRAM AND COMPUTER READABLE RECORDING MEDIUM WHICH RECORDS VOICE RECOGNITION PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To obtain a language model having high estimation precision and a voice recognition device having high recognition precision.

SOLUTION: A learning text data tree structure clustering means 2001 conducts a tree structure clustering to hierarchically divide learning text data 1001 so as to have a linguistically similar nature and generates a tree structure learning text data cluster 2002. A language model generating means 1004 generates a tree structure cluster language model 2003 employing each learning text data which belongs to the cluster 2002.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's]

***This Page Blank (uspto)***

decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

***This Page Blank (uspto)***

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2002-91484  
(P2002-91484A)

(43)公開日 平成14年3月27日(2002.3.27)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード(参考)
G 1 0 L 15/18		G 1 0 L 3/00	5 3 7 F 5 D 0 1 5

審査請求 未請求 請求項の数19 O L (全 19 頁)

(21)出願番号 特願2000-280655(P2000-280655)

(22)出願日 平成12年9月14日(2000.9.14)

(71)出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72)発明者 石井 純

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

(74)代理人 100066474

弁理士 田澤 博昭 (外1名)

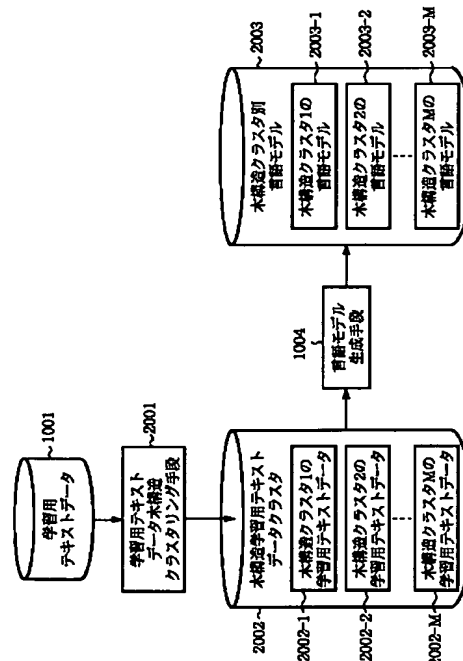
Fターム(参考) 5D015 AA01 GG01 GG05 HH23

(54)【発明の名称】 言語モデル生成装置及びこれを用いた音声認識装置、言語モデル生成方法及びこれを用いた音声認識方法、並びに言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体及

(57)【要約】

【課題】 推定精度の高い言語モデル、認識精度の高い音声認識装置を得る。

【解決手段】 学習用テキストデータ木構造クラスタリング手段2001は、学習用テキストデータ1001を言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタ2002を生成する。言語モデル生成手段1004は、木構造学習用テキストデータクラスタ2002に属する各学習用テキストデータを用いて、木構造クラスタ別言語モデル2003を生成する。



## 【特許請求の範囲】

【請求項1】 学習用テキストデータを入力して、単語列の生起確率を求める言語モデルを生成する言語モデル生成装置において、

上記学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する学習用テキストデータ木構造クラスタリング手段と、

上記木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する言語モデル生成手段とを備えたことを特徴とする言語モデル生成装置。

【請求項2】 木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する言語モデル補間手段を備えたことを特徴とする請求項1記載の言語モデル生成装置。

【請求項3】 認識対象音声を入力して音声認識を行い音声認識結果を出力する音声認識装置において、

上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手段と、

音声の音響的な観測値系列の確率を求める音響モデルと、

学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルと、

上記木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する言語モデル選択手段と、

上記言語モデル選択手段により選択された言語モデルと上記音響モデルを用いて、上記音声特徴量抽出手段が抽出した音声特徴量に対して照合を行い音声認識結果を出力する照合手段とを備えたことを特徴とする音声認識装置。

【請求項4】 言語モデル選択手段が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択することを特徴とする請求項3記載の音声認識装置。

【請求項5】 認識対象音声を入力して音声認識を行い音声認識結果を出力する音声認識装置において、

上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手段と、

音声の音響的な観測値系列の確率を求める音響モデルと、

学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルと、

上記木構造クラスタ別言語モデルから、音声認識結果候

補の単語列に対して生起確率の高い複数の言語モデルを選択する複数言語モデル選択手段と、

上記複数言語モデル選択手段によって選択された複数の言語モデルを入力して混合言語モデルを生成する混合言語モデル生成手段と、

上記混合言語モデル生成手段により生成された言語モデルと上記音響モデルを用いて、上記音声特徴量抽出手段が抽出した音声特徴量に対して照合を行い音声認識結果を出力する照合手段とを備えたことを特徴とする音声認識装置。

【請求項6】 複数言語モデル選択手段が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択することを特徴とする請求項5記載の音声認識装置。

【請求項7】 木構造クラスタ別言語モデルが、木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理が行われた補間処理された木構造クラスタ別言語モデルであることを特徴とする請求項3又は請求項5記載の音声認識装置。

【請求項8】 学習用テキストデータを入力して、単語列の生起確率を求める言語モデルを生成する言語モデル生成方法において、

上記学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する第1のステップと、

上記木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する第2のステップとを備えたことを特徴とする言語モデル生成方法。

【請求項9】 木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する第3のステップを備えたことを特徴とする請求項8記載の言語モデル生成方法。

【請求項10】 認識対象音声を入力して音声認識を行い音声認識結果を出力する音声認識方法において、上記認識対象音声を入力し音声特徴量を抽出する第1のステップと、

学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する第2のステップと、

音声の音響的な観測値系列の確率を求める音響モデルと、上記第2のステップで選択された言語モデルを用いて、上記第1のステップで抽出した音声特徴量に対して照合を行い音声認識結果を出力する第3のステップとを備えたことを特徴とする音声認識方法。

【請求項11】 第2のステップで、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択することを特徴とする請求項10記載の音声認識方法。

【請求項12】 認識対象音声を入力した音声認識を行い音声認識結果を出力する音声認識方法において、上記認識対象音声を入力し音声特徴量を抽出する第1のステップと、

学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率が高い複数の言語モデルを選択する第2のステップと、

上記第2のステップで選択された複数の言語モデルを入力して混合言語モデルを生成する第3のステップと、音声の音響的な観測値系列の確率を求める音響モデルと、上記第3のステップで生成された言語モデルを用いて、上記第1のステップで抽出した音声特徴量に対して照合を行い音声認識結果を出力する第4のステップとを備えたことを特徴とする音声認識方法。

【請求項13】 第2のステップで、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択することを特徴とする請求項12記載の音声認識方法。

【請求項14】 学習用テキストデータを入力して、単語列の生起確率を求める言語モデルを生成する言語モデル生成プログラムを記録した記録媒体であって、

上記学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する学習用テキストデータ木構造クラスタリング手順と、

上記木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する言語モデル生成手順とを実現させる言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項15】 木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する言語モデル補間手順を実現させる請求項14記載の言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項16】 認識対象音声を入力して音声認識を行い音声認識結果を出力する音声認識プログラムを記録した記録媒体であって、

上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手順と、

学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各

木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する言語モデル選択手順と、

音声の音響的な観測値系列の確率を求める音響モデルと、上記言語モデル選択手順により選択された言語モデルを用いて、上記音声特徴量抽出手順により抽出された音声特徴量に対して照合を行い音声認識結果を出力する照合手順とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項17】 言語モデル選択手順が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択することを特徴とする請求項16記載の音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項18】 認識対象音声を入力して音声認識を行い音声認識結果を出力する音声認識プログラムを記録した記録媒体であって、

上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手順と、

学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率の高い複数の言語モデルを選択する複数言語モデル選択手順と、

上記複数言語モデル選択手順によって選択された複数の言語モデルを入力して混合言語モデルを生成する混合言語モデル生成手順と、

音声の音響的な観測値系列の確率を求める音響モデルと、上記混合言語モデル生成手順により生成された言語モデルを用いて、上記音声特徴量抽出手順により抽出された音声特徴量に対して照合を行い音声認識結果を出力する照合手順とを実現させる音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。

【請求項19】 複数言語モデル選択手順が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択することを特徴とする請求項18記載の音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、音声認識を行う際に参照する言語モデル生成装置及びこれを用いた音声認識装置、言語モデル生成方法及びこれを用いた音声認識方法、並びに言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体に関するものである。

【0002】

10

20

30

40

50

【従来の技術】近年、使用話者が単語を連続して入力できる連続音声認識技術の実用化検討が盛んに行われている。連続音声認識は、単語の復号列が最大事後確率を持\*

$$\hat{W} = \arg \max_w P(O|W)P(W)$$

\*つように、音声の音響的な観測系列に基づいて復号することである。これは次の(1)式で表される。  
【数1】

(1)

ここで、Oは音声の音響的な観測値系列 $[o_1, o_2, o_3, \dots, o_T]$ であり、Wは単語列 $[w_1, w_2, w_3, \dots, w_n]$ である。P(O|W)は単語列Wが与えられたときの観測値系列Oに対する確率であり、音響モデルによって計算するものである。P

(W)は単語列Wの生起確率(出現確率)であり、言語モデルによって計算するものである。

【0003】音声認識については、森北出版(株)から出版されている古井貞熙著の「音声情報処理」(以降、文献1とする)、電子情報通信学会から出版されている中川聖一著の「確率モデルによる音声認識」(以降、文献2とする)、NTTアドバンステクノロジー(株)から出版されているLawrence Rabiner、B i i n g - H w a n g J u a n g 著、古井貞熙監訳の「音声認識の基礎(上、下)」(以降、文献3とする)に詳しく説明されている。

※【0004】音響モデルによって計算するP(O|W)は、最近統計的手法である隠れマルコフモデル(HMM)を用いる検討が盛んである。隠れマルコフモデルを用いた音響モデルは、例えば文献3の6章に詳しく述べられている。

【0005】また、言語モデルによって計算するP(W)は統計的な手法を用いることが多く、代表的なものにN-gramモデルがある(Nは2以上)。これらについては、東京大学出版会から出版されている北研二著の「確率的言語モデル」(以下文献4とする)の3章において詳しく説明されている。N-gramモデルは、直前の(N-1)個の単語から次の単語への遷移確率を統計的に与えるものである。N-gramモデルによる単語列 $w_1^L = w_1 \dots w_L$ の生起確率は、次の(2)式によって与えられる。

※

【数2】

$$P(w_1^L) = \prod_{t=1}^L P(w_t | w_{1:t-1}^L) \quad (2)$$

【0006】上記(2)式において、確率 $P(w_t | w_{1:t-1}^L)$ は(N-1)個の単語からなる単語列 $w_{1:t-1}^L$ の後に単語 $w_t$ が生起する確率であり、 $\prod$ は積を表している。例えば、「私・は・駅・へ・行く」★

$$P(\text{私} \cdot \text{は} \cdot \text{駅} \cdot \text{へ} \cdot \text{行く}) =$$

$$P(\text{私} | \#) P(\text{は} | \text{私}) P(\text{駅} | \text{は}) P(\text{へ} | \text{駅}) P(\text{行く} | \text{へ})$$

$$P(\# | \text{行く})$$

(3)

【0007】確率 $P(w_t | w_{1:t-1}^L)$ は学習用テキストデータの単語列の相対頻度によって求められる。単語列Wの学習用テキストデータにおける出現頻度をC(W)とすれば、例えば、「私・は」の2-gram確率

$$P(\text{は} | \text{私}) = C(\text{私} \cdot \text{は}) / C(\text{私})$$

(4)

【0008】しかしながら、N-gramモデルの確率値を単純に相対頻度によって推定すると、学習用テキストデータ中に出現しない単語組を0にしてしまうという大きな欠点がある(ゼロ頻度問題)。また、例えば学習用テキストデータ中に出現したとしても出現頻度の小さな単語列に対しては、統計的に信頼性のある確率値を推定するのが難しい(スパースネスの問題)。これらの問題に対処するために、通常はスムージングあるいは平滑化と呼ばれる手法を用いる。スムージングについては、上記文献4の3.3章にいくつかの手法が述べられているので、ここでは、具体的な説明は省略する。

【0009】言語モデルの学習には、音声認識の対象とする分野や場面・状況の文を学習用テキストデータとし

★(・は単語の区切りを表す)といった単語列の生起確率を2-gram(バイグラム)で求める場合は、次の(3)式ようになる。(3)式において、#は文頭、文末を表す記号である。

☆率 $P(\text{は} | \text{私})$ は、次の(4)式によって計算される。(4)式において、C(私・は)は単語列「私・は」の出現頻度、C(私)は「私」の出現頻度である。

て用いるが、実際のアプリケーションでは、音声認識の対象がさまざまな分野や、さまざまな場面・状況の音声である場合が多い。単語列の生起確率は分野、場面・状況が違えば異なった確率となるので、分野、場面・状況の異なりを無視して学習用テキストデータを一括して学習して言語モデルを生成した場合は、言語モデルの精度は良くない。

【0010】このような、さまざまな分野や、さまざまな場面・状況を音声認識の対象とした音声認識装置の性能を上げるために、言語モデルの学習用テキストデータをクラスタリングして、分割されたクラスタ毎に言語モデルを作成する方法が検討されている。従来技術としては、例えば、公開特許公報2000-75886号の



「統計的言語モデル生成装置及び音声認識装置」(以降、文献5とする)がある。ここで、クラスタとは、例えばクラスタ1が政治、クラスタ2がスポーツといった分野別の分割や、文の距離を定義して文をクラスタリングして得ることができる。

【0011】学習用テキストデータをクラスタに分割した場合には、クラスタ当たりの学習用テキストデータは少なくなるので、更に前述のゼロ頻度問題やスパースネスの問題が大きくなる。これに対して文献5では、クラスタに分割しない全学習用テキストデータを用いて推定した言語モデル $LM_k$ と、クラスタに分割された学習用テキストデータを用いて推定したクラスタ別の言語モデル $LM_k^k$ ( $k$ はクラスタ番号)を用いて、最大事後確率推定法によって $LM_{k..}$ を推定することで精度の高い言語モデルを得ている。

【0012】図13は文献5に記述されている従来の言語モデル生成装置の構成を示すブロック図である。図において、1001は言語モデルの学習用テキストデータ、1002は学習用テキストデータクラスタリング手段、1003は学習用テキストデータクラスタ、1003-1~1003-Mはクラスタ1~Mの学習用テキストデータ、1004は言語モデル生成手段、1005はクラスタ別言語モデル、1005-1~1005-Mはクラスタ1~Mの言語モデルである。

【0013】次に動作について説明する。学習用テキストデータ1001は、言語モデルを学習するためのテキストデータであり、音声認識装置が認識対象とする単語や文を文字にしたものである。この学習用テキストデータ1001は、学習用テキストデータクラスタリング手段1002へ入力される。

【0014】学習用テキストデータクラスタリング手段1002は、学習用テキストデータ1001をクラスタリングする。文献5では、 $k$ -means法に類似した方法を用いてテキストを文単位でクラスタリングしている。通常の $k$ -means法と異なる点は、(1)クラスタ中心ベクトルを、そのクラスタに属する文で生成される言語モデルとすること、(2)距離尺度に文の生成確率を用いていることである。また、言語モデルにはN-gramモデルを用いている。

【0015】学習用テキストデータクラスタ1003は、学習用テキストデータクラスタリング手段1002によって、M個のクラスタにクラスタリングされたクラスタ1の学習用テキストデータ1003-1~クラスタMの学習用テキストデータ1003-Mで構成されている。

【0016】言語モデル生成手段1004は、学習用テキストデータクラスタリング手段1002によって得られたクラスタ1の学習用テキストデータ1003-1~クラスタMの学習用テキストデータ1003-Mをそれぞれ入力して、クラスタ1の言語モデル1005-1~

クラスタMの言語モデル1005-Mで構成するクラスタ別言語モデル1005を生成する。言語モデル生成手段1004は、クラスタ毎の学習用テキストデータ数の減少による言語モデルの推定精度の低下を防ぐために、クラスタ分割しない全学習用テキストデータを用いて推定した言語モデル $LM_k$ と、クラスタに分割された学習用テキストデータを用いて推定したクラスタ別の言語モデル $LM_k^k$ を用いて、最大事後確率推定法によってクラスタ別の言語モデル $LM_{k..}$ を推定している。

【0017】次に上記言語モデル生成装置を用いた従来の音声認識装置の説明を行う。図14は文献5に開示された従来の音声認識装置の構成を示すブロック図である。図において、1101は認識対象音声、1102は音声特徴量抽出手段、1103は音響モデル、1104は言語モデル選択手段、1105は照合手段、1106は音声認識結果である。クラスタ別言語モデル1005は、図13と同一の機能ブロックであり、同一の符号を付すと共に説明は省略する。

【0018】次に動作について説明する。認識対象音声1101は認識対象とする音声であり、音声特徴量抽出手段1102へ入力される。音声特徴量抽出手段1102は、認識対象音声1101に含まれている音声特徴量を抽出する。音響モデル1103は音声の音響的な照合を行うためのモデルである。音響モデル1103は、例えば、多数の話者が発声した文や単語の音声を用いて学習した、前後音素環境を考慮した音素を認識ユニットとしたHMMを用いる。

【0019】言語モデル選択手段1104は、言語モデル生成装置を用いて生成したクラスタ1の言語モデル1005-1~クラスタMの言語モデル1005-Mで構成されるクラスタ別言語モデル1005の中から、照合手段1105で用いる言語モデルを選択する。文献5では、クラスタに分割する前の不特定言語モデルを用いて照合を行い、得られた認識結果候補の単語列に対して、最も生起確率が高いクラスタ別言語モデルを、クラスタ1の言語モデル1005-1~クラスタMの言語モデル1005-Mから1つ選択している。

【0020】照合手段1105は、言語モデル選択手段1104によって選択された言語モデルが設定している認識対象の単語 $[W(1), W(2), \dots, W(w_n)]$ ( $w_n$ は認識対象とする単語数)の発音表記を認識ユニットラベル表記に変換し、このラベルにしたがって、音響モデル1103に格納されている音素単位のHMMを連結し、認識対象単語の標準パターン $[\lambda_{w(1)}, \lambda_{w(2)}, \dots, \lambda_{w(w_n)}]$ を作成する。

【0021】そして、照合手段1105は、認識対象単語の標準パターンと選択された言語モデルによって表される単語列の生起確率を用いて、音声特徴量分析手段1102の出力である音声特徴量に対して照合を行い、音声認識結果1106を出力する。音声認識結果1106

は、認識対象音声1101に対して、認識対象単語で最も照合スコアが高い単語の単語番号系列 $R_n = [r(1), r(2), \dots, r(m)]$ を計算し、単語番号に対応する単語 $R_w = [W(r(1)), W(r(2)), \dots, W(r(m))]$ を出力する。ここで、 $r(i)$ は音声認識結果1106の単語系列の*i*番目の単語の単語番号を示す。また、*m*は認識単語系列の単語数を示す。

【0022】

【発明が解決しようとする課題】従来の言語モデル生成装置は以上のように構成されているので、クラスタリングによって分割するクラスタ数が増えると、クラスタ当たりの学習用テキストデータ数が少なくなり、言語モデルの推定精度が悪くなるので音声認識精度が高くないという課題があった。

【0023】また、分割するクラスタ数が増えると、1発声が複数のクラスタの言語性質を持つような場合、認識率が高くないという課題があった。

【0024】この発明は、上記のような課題を解決するためになされたものであり、推定精度の高い言語モデルを作成できる言語モデル生成装置、言語モデル生成方法及び言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体を得ることを目的とする。

【0025】また、この発明は、推定精度の高い言語モデルを用いて、音声認識精度の高い音声認識装置、音声認識方法及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体を得ることを目的とする。

【0026】

【課題を解決するための手段】この発明に係る言語モデル生成装置は、学習用テキストデータを入力して、単語列の生起確率を求める言語モデルを生成するものにおいて、上記学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する学習用テキストデータ木構造クラスタリング手段と、上記木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する言語モデル生成手段とを備えたものである。

【0027】この発明に係る言語モデル生成装置は、木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する言語モデル補間手段を備えたものである。

【0028】この発明に係る音声認識装置は、認識対象音声を入力して音声認識を行い音声認識結果を出力するものにおいて、上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手段と、音声の音響的な観測値系列の確率を求める音響モデルと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割す

る木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルと、上記木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する言語モデル選択手段と、上記言語モデル選択手段により選択された言語モデルと上記音響モデルを用いて、上記音声特徴量抽出手段が抽出した音声特徴量に対して照合を行い音声認識結果を出力する照合手段とを備えたものである。

【0029】この発明に係る音声認識装置は、言語モデル選択手段が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択するものである。

【0030】この発明に係る音声認識装置は、認識対象音声を入力して音声認識を行い音声認識結果を出力するものにおいて、上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手段と、音声の音響的な観測値系列の確率を求める音響モデルと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルと、上記木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率の高い複数の言語モデルを選択する複数言語モデル選択手段と、上記複数言語モデル選択手段によって選択された複数の言語モデルを入力して混合言語モデルを生成する混合言語モデル生成手段と、上記混合言語モデル生成手段により生成された言語モデルと上記音響モデルを用いて、上記音声特徴量抽出手段が抽出した音声特徴量に対して照合を行い音声認識結果を出力する照合手段とを備えたものである。

【0031】この発明に係る音声認識装置は、複数言語モデル選択手段が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択するものである。

【0032】この発明に係る音声認識装置は、木構造クラスタ別言語モデルが、木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理が行われた補間処理された木構造クラスタ別言語モデルであることを特徴とするものである。

【0033】この発明に係る言語モデル生成方法は、学習用テキストデータを入力して、単語列の生起確率を求める言語モデルを生成するものにおいて、上記学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する第1のステップと、上記木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する第2のステップとを備えたものである。

【0034】この発明に係る言語モデル生成方法は、木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する第3のステップを備えたものである。

【0035】この発明に係る音声認識方法は、認識対象音声を入力して音声認識を行い音声認識結果を出力するものにおいて、上記認識対象音声を入力し音声特徴量を抽出する第1のステップと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する第2のステップと、音声の音響的な観測値系列の確率を求める音響モデルと、上記第2のステップで選択された言語モデルを用いて、上記第1のステップで抽出した音声特徴量に対して照合を行い音声認識結果を出力する第3のステップとを備えたものである。

【0036】この発明に係る音声認識方法は、第2のステップで、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択するものである。

【0037】この発明に係る音声認識方法は、認識対象音声を入力した音声認識を行い音声認識結果を出力するものにおいて、上記認識対象音声を入力し音声特徴量を抽出する第1のステップと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率が高い複数の言語モデルを選択する第2のステップと、上記第2のステップで選択された複数の言語モデルを入力して混合言語モデルを生成する第3のステップと、音声の音響的な観測値系列の確率を求める音響モデルと、上記第3のステップで生成された言語モデルを用いて、上記第1のステップで抽出した音声特徴量に対して照合を行い音声認識結果を出力する第4のステップとを備えたものである。

【0038】この発明に係る音声認識方法は、第2のステップで、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択するものである。

【0039】この発明に係る言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体は、学習用テキストデータを入力して、単語列の生起確率を求める言語モデルを生成するものであって、上記学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する学習用テキストデ

ータ木構造クラスタリング手順と、上記木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する言語モデル生成手順とを実現させるものである。

【0040】この発明に係る言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体は、木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する言語モデル補間手順を実現させるものである。

【0041】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、認識対象音声を入力して音声認識を行い音声認識結果を出力するもので、上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手順と、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する言語モデル選択手順と、音声の音響的な観測値系列の確率を求める音響モデルと、上記言語モデル選択手順により選択された言語モデルを用いて、上記音声特徴量抽出手順により抽出された音声特徴量に対して照合を行い音声認識結果を出力する照合手順とを実現させるものである。

【0042】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、言語モデル選択手順が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択するものである。

【0043】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、認識対象音声を入力して音声認識を行い音声認識結果を出力するものであって、上記認識対象音声を入力し音声特徴量を抽出する音声特徴量抽出手順と、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率の高い複数の言語モデルを選択する複数言語モデル選択手順と、上記複数言語モデル選択手順によって選択された複数の言語モデルを入力して混合言語モデルを生成する混合言語モデル生成手順と、音声の音響的な観測値系列の確率を求める音響モデルと、上記混合言語モデル生成手順により生成された言語モデルを用いて、上記音声特徴量抽出手順により抽出された音声特徴量に対して照合を行い音声認識結果を出力する照合手順とを実現させるものである。

【0044】この発明に係る音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体は、複数言語

モデル選択手順が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択するものである。

【0045】

【発明の実施の形態】以下、この発明の一形態を説明する。

実施の形態1. 図1はこの発明の実施の形態1による言語モデル生成装置の構成を示すブロック図である。図において、2001は学習用テキストデータ木構造クラスタリング手段、2002は木構造学習用テキストデータクラスタ、2002-1~2002-Mは木構造クラスタ1~Mの学習用テキストデータ、2003は木構造クラスタ別言語モデル、2003-1~2003-Mは木構造クラスタ1~Mの言語モデルである。従来の言語モデル生成装置の構成を示す図13と同一の機能ブロックについては、同一の符号を付し説明を省略する。

【0046】なお、言語モデルの学習用テキストデータ1001は、音声認識の認識対象とする分野や場面・状況において用いられる単語や文を文字化したものである。例えば、アナウンサーが発声する政治のニュースを音声認識対象とした場合は、新聞の政治欄の記事や、政治の放送ニュースの発声内容を文字として書き起こしたテキストデータである。

【0047】次に動作について説明する。図2はこの発明の実施の形態1による言語モデル生成装置における言語モデル生成方法を示すフローチャートである。学習用テキストデータ木構造クラスタリング手段2001は、ステップST101において、学習用テキストデータ1001を入力し、ステップST102において、クラスタリングの階層Iを0とし、ステップST103において、初めに学習用テキストデータ1001の全てに対してクラスタリングを行う。この学習用テキストデータ1001全てに対するクラスタリングを、階層0のクラスタリングとする。

【0048】ここで、クラスタリングとは、人手で2つ以上の分野に分けることや、文献5に示してあるk-meansアルゴリズムに類似した方法を用いて、学習用テキストデータを2つ以上の集合に分割することである。クラスタリングによって得られるクラスタに属する学習用テキストデータは、言語的に類似した性質を持つものとなる。

【0049】図3は学習用テキストデータ木構造クラスタリング手段2001で行われる学習用テキストデータ木構造クラスタリングの説明図であり、文を単位として階層的にクラスタリングしている様子を示したものである。図3では、階層0の木構造クラスタ00のクラスタリングにより、学習用テキストデータ1001全てを2つのクラスタに分割している。分割された学習用テキストデータの集合は、階層1の木構造クラスタ10と木構造クラスタ11となっている。

【0050】図2のステップST104において、学習用テキストデータ木構造クラスタリング手段2001

は、階層Iをインクリメントし、ステップST105において、学習用テキストデータ木構造クラスタリング手段2001は、階層I-1（ここでは、I=0）でクラスタリングされた各クラスタに属する学習用テキストデータに対してクラスタリングを行う。図3では、階層1のクラスタリングにより、階層1の木構造クラスタ10から階層2の木構造クラスタ20と木構造クラスタ21を生成し、木構造クラスタ11から階層2の木構造クラスタ22と木構造クラスタ23を生成している。

【0051】ステップST106において、クラスタ数が予め定めた数Mになったかを調べて、予め定めた数Mにならない場合には、ステップST104に戻り、階層Iをインクリメントし、ステップST105のクラスタリングの処理を繰り返す。以上の処理をクラスタ数が予め定めた数Mになるまで繰り返して、木構造クラスタ1の学習用テキストデータ2002-1~木構造クラスタMの学習用テキストデータ2002-Mを生成する。

【0052】予め定めたクラスタ数まで学習用テキストデータの木構造クラスタリングを行った後に、ステップST107において、言語モデル生成手段1004は、クラスタリングされた木構造クラスター別に、各クラスタに属する学習用テキストデータを用いて言語モデルの生成を行い、木構造クラスタ1の言語モデル2003-1~木構造クラスタMの言語モデル2003-Mで構成される木構造クラスター別言語モデル2003を生成する。

【0053】上記ステップST106において、階層的な学習用テキストデータのクラスタリングを、予め定めたクラスタ数Mになるまで繰り返す。ここでは、クラスタ数をクラスタリングの終了の基準にしているが、階層数を基準としても、クラスタ内の学習用テキストデータ数がある値以下であるならば、クラスタリングを終了するとしても良い。階層的なクラスタリングによって得られるクラスタは、階層が下になるほどクラスタに属する学習用テキストデータの性質は分野や場面・状況の違いをよく表現している。

【0054】図4は木構造クラスター別の言語モデル生成の説明図である。図4では木構造のノードが学習用テキストデータの木構造クラスタを表しており、各木構造クラスタ毎にそこに属する学習用テキストデータを用いて言語モデルの生成を行う。木構造の親ノードの木構造クラスタは、子ノードの木構造クラスタに属する学習用テキストデータ全てを含むものとなっている。図4では、例えば、木構造クラスタ00に属する学習用テキストデータを用いて生成した言語モデルが、木構造クラスタ00の言語モデルLM00、木構造クラスタ10に属する学習用テキストデータを用いて生成した言語モデルが、木構造クラスタ10の言語モデルLM10にそれぞれ対応している。

【0055】生成される言語モデルの性質は、下層の木構造クラスタの言語モデルへいくほど、分野や場面・状況の違いによる言語の性質の違いを、より表現した言語モデルとなる。また、上層の木構造クラスタの言語モデルは、分野や場面・状況の違いによる言語の性質の違いは細かく表していないが、複数の分野や場面・状況の言語特徴を含んでいるので、発声が複数の分野や場面・状況を含んでいる場合には、有効な言語モデルとなっている。さらに、上層の木構造クラスタの言語モデルは学習テキストデータが多いので、木構造クラスタと同数のクラスタ数に一度に分割した場合に比べてパラメータ推定精度が高い。

【0056】言語モデルの生成の具体的方法は、文献4の3章から5章に述べられている、N-gramモデル、隠れマルコフモデル、確率文脈自由文法等である。

【0057】また、この実施の形態1における言語モデル生成方法を言語モデル生成プログラムとして記録媒体に記録することもできる。この場合には、学習用テキストデータ木構造クラスタリング手段2001と同様の処理を実現する学習用テキストデータ木構造クラスタリング手順と、言語モデル生成手段1004と同様の処理を実現する言語モデル生成手順とから構成される言語モデル生成プログラムを記録媒体に記録する。

【0058】以上のように、この実施の形態1の言語モデル生成装置及び言語モデル生成方法によれば、学習用テキストデータを階層的に木構造クラスタリングし、各木構造クラスタに属する学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成するので、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、認識率の高い言語モデルが生成できる効果が得られる。また、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高\*

$$P'_i(w_n | w_{n+1-N}^{n-1}) = \sum_{S \in \Omega} \alpha_i P_i(w_n | w_{n+1-N}^{n-1}) \quad (5)$$

【0063】上記(5)式において、 $P'_i(w_n | w_{n+1-N}^{n-1})$ は補間処理された木構造クラスタSの言語モデルにおける単語列 $w_{n+1-N}^{n-1}$ に続いて $w_n$ が生起する確率、 $\Omega$ は木構造クラスタSとその親ノードのクラスタ番号の集合、 $P_i(w_n | w_{n+1-N}^{n-1})$ は木構造クラスタiの言語モデルにおける単語列 $w_{n+1-N}^{n-1}$ に続いて $w_n$ が生起する確率、 $\alpha_i$ は重み係数である。この $\alpha_i$ は、例えば、文献4の3章に述べられている削除補間法によって推定可能である。

【0064】この説明では、 $P_i(w_n | w_{n+1-N}^{n-1})$ は補間する前の生起確率としたが、木構造の上層から補間し、補間処理された生起確率 $P'_i(w_n | w_{n+1-N}^{n-1})$ を用いても良い。木構造クラスタでは、下層のクラスタは学習用テキストデータが少量であるので、言語

\*い言語モデルが生成できる効果が得られる。

【0059】実施の形態2. 図5はこの発明の実施の形態2による言語モデル生成装置の構成を示すブロック図である。図において、3001は言語モデル補間手段、3002は補間処理された木構造クラスタ別言語モデル、3002-1~3002-Mは補間処理された木構造クラスタ1~Mの言語モデルである。実施の形態1の図1と同一の機能ブロックについては、同一の符号を付し説明を省略する。

10 【0060】次に動作について説明する。図6はこの発明の実施の形態2による言語モデル生成装置における言語モデル生成方法を示すフローチャートである。ステップST201からステップST207までの処理は、実施の形態1の図2におけるステップST101からステップST107までの処理と同一である。

【0061】ステップST208において、言語モデル補間手段3001は、言語モデル生成手段1004によって生成された木構造クラスタ別言語モデルである木構造クラスタ1の言語モデル2003-1~木構造クラスタMの言語モデル2003-Mを入力し、補間処理された木構造クラスタ1の言語モデル3002-1~補間処理された木構造クラスタMの言語モデル3002-Mを生成する。このときの補間処理は、補間対象のクラスタ言語モデルが位置する木構造のノードの親ノードの木構造クラスタの言語モデルを用いて補間処理を行う。

【0062】図4の例では、木構造クラスタ20の言語モデルLM20を補間する場合は、親ノードである木構造クラスタ10の言語モデルLM10と、更に上層の親ノードである木構造クラスタ00の言語モデルLM00とを用いて補間する。この補間処理において、例えば言語モデルがN-gramモデルである場合には、単語列 $w_{n+1-N}^{n-1}$ に続いて $w_n$ が生起する確率がパラメータであり、次の(5)式によって求める。

【数3】

$$P'_i(w_n | w_{n+1-N}^{n-1}) = \sum_{S \in \Omega} \alpha_i P_i(w_n | w_{n+1-N}^{n-1}) \quad (5)$$

モデル生成において、ゼロ頻度問題やスパースネスの問題が生じやすいが、このように、学習用テキストデータ数が多い親ノードのクラスタの言語モデルを用いて、パラメータすなわち単語列 $w_{n+1-N}^{n-1}$ に続いて $w_n$ が生起する確率の補間処理を行うので、言語モデル推定精度が高くなる。

【0065】また、実施の形態2における言語モデル生成方法を言語モデル生成プログラムとして記録媒体に記録することもできる。この場合には、学習用テキストデータ木構造クラスタリング手段2001と同様の処理を実現する学習用テキストデータ木構造クラスタリング手順と、言語モデル生成手段1004と同様の処理を実現する言語モデル生成手順と、言語モデル補間手段3001と同様の処理を実現する言語モデル補間手順とから構

成される言語モデル生成プログラムを記録媒体に記録する。

【0066】以上のように、この実施の形態2の言語モデル生成装置及び言語モデル生成方法によれば、学習用テキストデータを階層的に木構造クラスタリングし、各木構造クラスタに属する学習用テキストデータを用いて木構造クラスタ別言語モデルを生成し、生成されたクラスタ言語モデルを木構造の親ノードのクラスタ言語モデルを用いて補間するので、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、さらに認識率の高い言語モデルを生成できるという効果が得られる。

【0067】また、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できるという効果が得られる。

【0068】実施の形態3. 図7はこの発明の実施の形態3による音声認識装置の構成を示すブロック図である。図において、実施の形態1の図1、及び従来の音声認識装置の図14と同一の機能ブロックについては、同一の符号を付し説明を省略する。

【0069】次に動作について説明する。図8はこの発明の実施の形態3による音声認識装置における音声認識方法を示すフローチャートである。音声特徴量抽出手段1102は、ステップST301において認識対象音声1101を入力し、ステップST302において音声特徴量を抽出する。ここで、音声特徴量とは少ない情報量で音声の特徴を表すものであり、例えば、文献1の5章で述べているようなケプストラム、ケプストラムの動的

特徴で構成する特徴ベクトルである。

【0070】ステップST303において、言語モデル選択手段1104は、照合手段1105で用いる言語モデルを、木構造クラスタ別言語モデル2003の木構造クラスタ1の言語モデル2003-1～木構造クラスタMの言語モデル2003-Mから1つ選択する。言語モデルの選択は、例えば文献5に示されている方法を用い、最も生起確率が高い木構造クラスタの言語モデルを選択する。

【0071】ステップST304において、照合手段1105は、言語モデル選択手段1104によって選択された木構造クラスタ言語モデルと、音響モデル1103を入力して認識対象音声1101の音声特徴量に対して照合を行い、最も尤度（照合スコア）が高い単語列を音声認識結果1106として出力する。

【0072】この場合の照合処理を具体的に説明する。照合手段1105は、言語モデル選択手段1104によって選択された木構造クラスタ言語モデルが設定している認識対象の単語 $[W(1), W(2), \dots, W(w_n)]$ （ $w_n$ は認識対象とする単語数）の発音表記

を、認識ユニットラベル表記に変換し、このラベルにしたがって、音響モデル1103に格納されている音素ユニットのHMMを連結し、認識対象単語の標準パターン $[\lambda_{w(1)}, \lambda_{w(2)}, \dots, \lambda_{w(w_n)}]$ を作成する。

【0073】そして、照合手段1105は、認識対象単語標準パターンと選択された木構造クラスタ言語モデルによって表される単語列の生起確率を用いて、音声特徴量分析手段1102の出力である音声特徴量に対して照合を行い、音声認識結果1106を出力する。音声認識結果1106は、認識対象音声に対して認識対象単語で最も尤度が高い単語の単語番号系列 $R_n = [r(1), r(2), \dots, r(m)]$ を計算し、単語番号に対応する単語 $R_w = [W(r(1)), W(r(2)), \dots, W(r(m))]$ を出力する。ここで、 $r(i)$ は音声認識結果の単語系列の*i*番目の単語の単語番号を示し、*m*は認識単語系列の単語数を示す。

【0074】以上は、選択対象の木構造クラスタ別言語モデルを、実施の形態1で生成した木構造クラスタ1の言語モデル2003-1～木構造クラスタMの言語モデル2003-Mとして説明したが、実施の形態2で生成した補間処理された木構造クラスタ1の言語モデル3002-1～補間処理された木構造クラスタMの言語モデル3002-Mとしても良い。

【0075】また、実施の形態3における音声認識方法を音声認識プログラムとして記録媒体に記録することもできる。この場合には、実施の形態1の言語モデル生成プログラムに加えて、音声特徴量抽出手段1102と同様の処理を実現する音声特徴量抽出手順と、言語モデル選択手段1104と同様の処理を実現する言語モデル選択手順と、照合手段1105と同様の処理を実現する照合手順を含む音声認識プログラムを記録媒体に記録する。

【0076】以上のように、この実施の形態3における音声認識装置及び音声認識方法によれば、学習用テキストデータ1001を階層的に木構造クラスタリングし、各木構造クラスタに属する学習用テキストデータ2002-1～2002-Mを用いて、木構造クラスタ別言語モデル2003-1～2003-Mを生成するので、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、この木構造クラスタ別言語モデル2003から言語モデルを選択して音声認識を行うので、認識精度が高い音声認識ができるという効果が得られる。

【0077】また、認識対象の音声は複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した木構造クラスタ言語モデルを選択し音声認識を行うので、認識性能が高い音声認識ができる効果が得られる。

【0078】実施の形態4. 図9はこの発明の実施の形態4による音声認識装置の構成を示すブロック図であ

10

20

30

40

50

る。図において、5001は複数言語モデル選択手段、5002は混合言語モデル生成手段である。実施の形態3の図7と同一の機能ブロックについては、同一の符号を付し説明を省略する。

【0079】次に動作について説明する。図10はこの発明の実施の形態4による音声認識装置における音声認識方法を示すフローチャートである。ステップST401及びステップST402の処理は、実施の形態3における図8のステップST301及びステップST302の処理と同一である。

【0080】ステップST403において、複数言語モデル選択手段5001は、木構造クラスタ1の言語モデル

$$P_m(w_n | w_{n+1-N}^{n-1}) = \sum_{i=1}^M \beta_i P_i(w_n | w_{n+1-N}^{n-1}) \quad (6)$$

【0082】上記(6)式において、 $P_m(w_n | w_{n+1-N}^{n-1})$ は混合言語モデルの生起確率であり、 $\Psi$ は複数言語モデル選択手段5001によって選択された木構造クラスタ言語モデルの番号の集合、 $P_i(w_n | w_{n+1-N}^{n-1})$ は選択された言語モデルの生起確率であり、 $\beta_i$ は重み係数である。ここで $\beta_i$ については、例えば文献5に示されている言語モデル選択時の生起確率にしたがって、生起確率が高い言語モデルは $\beta_i$ が大きくなるように設定する。

【0083】ステップST405において、照合手段1105は、混合言語モデル生成手段5002によって生成された混合言語モデルと、音響モデル1103を入力し、認識対象音声1101の音声特徴量に対して照合を行い、最も尤度が高い単語列を音声認識結果1106として出力する。

【0084】以上は、選択対象の木構造クラスタ言語モデルを、実施の形態1で生成した木構造クラスタ1の言語モデル2003-1～木構造クラスタMの言語モデル2003-Mとして説明したが、実施の形態2で生成した補間処理された木構造クラスタ1の言語モデル3002-1～補間処理された木構造クラスタMの言語モデル3002-Mとしても良い。

【0085】また、実施の形態4における音声認識方法を音声認識プログラムとして記録媒体に記録することもできる。この場合には、実施の形態1の言語モデル生成プログラムに加えて、音声特徴量抽出手段1102と同様の処理を実現する音声特徴量抽出手順と、照合手段1105と同様の処理を実現する照合手順と、複数言語モデル選択手段5001と同様の処理を実現する複数言語モデル選択手順と、混合言語モデル生成手段5002と同様の処理を実現する混合言語モデル生成手順とを含む音声認識プログラムを記録媒体に記録する。

【0086】以上のように、この実施の形態4における音声認識装置及び音声認識方法によれば、学習用テキストデータ1001を階層的に木構造クラスタリングし、

\*ル2003-1～木構造クラスタMの言語モデル2003-Mから2つ以上(K個以下)の木構造クラスタの言語モデルを選択する。言語モデルの選択は、例えば文献5に示されている方法を拡張し、生起確率が高い順からK個の言語モデルを選択する方法を用いる。

【0081】ステップST404において、混合言語モデル生成手段5002は、複数言語モデル選択手段5001によって選択された複数の木構造クラスタ言語モデルを入力し、1つの混合言語モデルを生成する。混合モデルは、例えばN-gramモデルであるならば、次の(6)式によって生起確率を計算する。

【数4】

各木構造クラスタの学習用テキストデータ2002-1～2002-Mを用いて、木構造クラスタ別言語モデル2003-1～2003-Mを生成し、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、この木構造クラスタ別言語モデル2003から複数選択した木構造クラスタ言語モデルによって混合言語モデルを生成して、音声認識に用いるので、さらに認識精度が高い音声認識ができるという効果が得られる。

【0087】また、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルを選択し混合言語モデルを生成して音声認識に用いるので、認識性能が高い音声認識ができる効果が得られる。

【0088】実施の形態5。図11はこの発明の実施の形態5による音声認識装置の構成を示すブロック図である。図において、6001は葉ノードのクラスタ別言語モデル、6001-1～6001-Lは葉ノードクラスタ1～Lの言語モデルである。実施の形態3の図7と同一の機能ブロックについては、同一の符号を付し説明を省略する。

【0089】次に動作について説明する。図12はこの発明の実施の形態5による音声認識装置における音声認識方法を示すフローチャートである。ステップST501及びステップST502の処理は、実施の形態3における図8のステップST301及びステップST302の処理と同一である。

【0090】ステップST503において、言語モデル選択手段1104は、木構造クラスタの葉ノードクラスタの言語モデルから、照合手段1105で用いる言語モデルを、葉ノードクラスタ1の言語モデル6001-1～葉ノードクラスタLの言語モデル6001-Lから1つ選択する。ここで、葉ノードクラスタの言語モデルとは、木構造の最も下層の木構造クラスタの言語モデルである。図4の例では、木構造クラスタ20の言語モデル



LM20、木構造クラスタ21の言語モデルLM21、木構造クラスタ22の言語モデルLM22、木構造クラスタ23の言語モデルLM23が葉ノードクラスタの言語モデルに相当する。

【0091】このような葉ノードクラスタの言語モデルは、分野や場面・状況の違いによる言語の性質の違いを詳細に表現するモデルとなっているので、分野や場面・状況が明確に分かれるような認識対象の音声である場合は有効である。また、全ての木構造クラスタ別の言語モデルを用いる場合に比べて、選択対象のクラスタ言語モデルの数が少ないので、省メモリ、演算量削減の効果がある。葉ノードクラスタの言語モデルの選択は、例えば文献5に示されている方法を用い、最も生起確率が高い葉ノードクラスタの言語モデルを選択する。

【0092】ステップST504において、照合手段1105は、言語モデル選択手段1104によって選択された葉ノードクラスタの言語モデルと、音響モデル1103を入力して、認識対象音声1101の音声特徴量に対して照合を行い、最も尤度が高い単語列を音声認識結果1106として出力する。

【0093】以上は、選択対象の葉ノードクラスタの言語モデルを、実施の形態1で生成した木構造クラスタ別言語モデル2003の葉ノードクラスタの言語モデルとしたが、実施の形態2で生成した補間処理された木構造クラスタ別言語モデル3002の葉ノードクラスタの言語モデルとしても良い。また、言語モデル選択手段1104を複数言語モデル選択手段5001とし、後段に混合言語モデル生成手段5002を接続し、混合言語モデルを用いて照合処理を行っても良い。

【0094】また、実施の形態5における音声認識方法を音声認識プログラムとして記録媒体に記録することもできる。この場合には、実施の形態1の言語モデル生成プログラムに加えて、音声特徴量抽出手段1102と同様の処理を実現する音声特徴量抽出手順と、言語モデル選択手段1104と同様の処理を実現する言語モデル選択手順と、照合手段1105と同様の処理を実現する照合手順を含む音声認識プログラムを記録媒体に記録する。

【0095】以上のように、この実施の形態5における音声認識装置及び音声認識方法によれば、学習用テキストデータ1001を階層的に木構造クラスタリングし、各木構造クラスタの学習用テキストデータ2002-1～2002-Mを用いて、木構造クラスタ言語モデル2003を生成するので、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、この木構造クラスタ言語モデル2003の葉ノードクラスタの言語モデル6001から選択した言語モデルを音声認識に用いるので、認識精度が高い音声認識ができると共に、言語モデルのメモリ容量を削減でき、言語モデルを選択する際の演算量

を削減できるという効果が得られる。

【0096】また、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の葉ノードクラスタの言語モデルを選択し混合言語モデルを生成すれば、複数の分野や場面・状況の言語特徴を学習した言語モデルを音声認識に用いることになるので、認識性能が高い音声認識ができる効果が得られる。

【0097】

【発明の効果】以上のように、この発明によれば、言語モデル生成装置が、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する学習用テキストデータ木構造クラスタリング手段と、木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する言語モデル生成手段とを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、認識率の高い言語モデルが生成できると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できる効果がある。

【0098】この発明によれば、言語モデル生成装置が、木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する言語モデル補間手段を備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、さらに認識率の高い言語モデルを生成できると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できるという効果がある。

【0099】この発明によれば、音声認識装置が、音声特徴量抽出手段と、音響モデルと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルと、木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する言語モデル選択手段と、選択された言語モデルと音響モデルを用いて、音声特徴量抽出手段が抽出した音声特徴量に対して照合を行い音声認識結果を出力する照合手段とを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、木構造クラスタ別言語モデルから言語モデルを選択して音声認識を行うので、認識精度が高い音声認識ができると共



に、認識対象の音声複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した木構造クラスタ言語モデルを選択し音声認識を行うので、認識性能が高い音声認識ができる効果がある。

【0100】この発明によれば、音声認識装置の言語モデル選択手段が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択することにより、言語モデルのメモリ容量を削減でき、言語モデルを選択する際の演算量を削減できるという効果がある。

【0101】この発明によれば、音声認識装置が、音声特徴量抽出手段と、音響モデルと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルと、木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率の高い複数の言語モデルを選択する複数言語モデル選択手段と、選択された複数の言語モデルを入力して混合言語モデルを生成する混合言語モデル生成手段と、生成された言語モデルと音響モデルを用いて、音声特徴量抽出手段が抽出した音声特徴量に対して照合を行い音声認識結果を出力する照合手段とを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、木構造クラスタ別言語モデルから複数選択した木構造クラスタ言語モデルによって混合言語モデルを生成して、音声認識に用いるので、さらに認識精度が高い音声認識ができると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルを選択し混合言語モデルを生成して音声認識に用いるので、認識性能が高い音声認識ができる効果がある。

【0102】この発明によれば、音声認識装置の複数言語モデル選択手段が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択することにより、言語モデルのメモリ容量を削減でき、言語モデルを選択する際の演算量を削減できるという効果がある。

【0103】この発明によれば、音声認識装置の木構造クラスタ別言語モデルが、木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理が行われた補間処理された木構造クラスタ別言語モデルであることにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、さらに認識率の高い言語モデルを生成できると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できるという効果がある。

【0104】この発明によれば、言語モデル生成方法として、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する第1のステップと、木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する第2のステップとを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、認識率の高い言語モデルが生成できると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できる効果がある。

【0105】この発明によれば、言語モデル生成方法として、木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する第3のステップを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、さらに認識率の高い言語モデルを生成できると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できるという効果がある。

【0106】この発明によれば、音声認識方法として、音声特徴量を抽出する第1のステップと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する第2のステップと、音響モデルと選択された言語モデルを用いて、音声特徴量に対して照合を行い音声認識結果を出力する第3のステップとを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、木構造クラスタ別言語モデルから言語モデルを選択して音声認識を行うので、認識精度が高い音声認識ができると共に、認識対象の音声複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した木構造クラスタ言語モデルを選択し音声認識を行うので、認識性能が高い音声認識ができる効果がある。

【0107】この発明によれば、音声認識方法の第2のステップで、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択することにより、言語モデルを選択する際の演算量を削減できるという効果がある。

【0108】この発明によれば、音声認識方法として、

音声特徴量を抽出する第1のステップと、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率が高い複数の言語モデルを選択する第2のステップと、選択された複数の言語モデルを入力して混合言語モデルを生成する第3のステップと、音響モデルと生成された言語モデルを用いて、抽出した音声特徴量に対して照合を行い音声認識結果を出力する第4のステップとを備えたことにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、木構造クラスタ別言語モデルから複数選択した木構造クラスタ言語モデルによって混合言語モデルを生成して、音声認識に用いるので、さらに認識精度が高い音声認識ができると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルを選択し混合言語モデルを生成して音声認識に用いるので、認識性能が高い音声認識ができる効果がある。

【0109】この発明によれば、音声認識方法の第2のステップで、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択することにより、言語モデルを選択する際の演算量を削減できるという効果がある。

【0110】この発明によれば、言語モデル生成プログラムを記録した記録媒体で、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、木構造学習用テキストデータクラスタを生成する学習用テキストデータ木構造クラスタリング手順と、木構造学習用テキストデータクラスタに属する各学習用テキストデータを用いて、木構造クラスタ別言語モデルを生成する言語モデル生成手順とを実現させることにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、認識率の高い言語モデルが生成できると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できる効果がある。

【0111】この発明によれば、言語モデル生成プログラムを記録した記録媒体で、木構造クラスタ別言語モデルが位置する木構造の上位に位置する木構造クラスタ別言語モデルを用いて補間処理を行い、補間処理された木構造クラスタ別言語モデルを生成する言語モデル補間手順を実現させることにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、さらに認識率の高い言語モデルを生成できると共に、認識対象の1発声が複

数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルが存在するので、認識率の高い言語モデルが生成できるという効果がある。

【0112】この発明によれば、音声認識プログラムを記録した記録媒体で、音声特徴量を抽出する音声特徴量抽出手順と、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して最も生起確率が高い言語モデルを選択する言語モデル選択手順と、音響モデルと選択された言語モデルを用いて、抽出された音声特徴量に対して照合を行い音声認識結果を出力する照合手順とを実現させることにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、木構造クラスタ別言語モデルから言語モデルを選択して音声認識を行うので、認識精度が高い音声認識ができると共に、認識対象の音声は複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した木構造クラスタ言語モデルを選択し音声認識を行うので、認識性能が高い音声認識ができる効果がある。

【0113】この発明によれば、音声認識プログラムの言語モデル選択手順が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから言語モデルを選択することにより、言語モデルを選択する際の演算量を削減できるという効果が得られる。

【0114】この発明によれば、音声認識プログラムを記録した記録媒体で、音声特徴量を抽出する音声特徴量抽出手順と、学習用テキストデータを言語的に類似した性質を持つように階層的に分割する木構造クラスタリングを行い、各木構造クラスタの学習用テキストデータを用いて生成された木構造クラスタ別言語モデルから、音声認識結果候補の単語列に対して生起確率の高い複数の言語モデルを選択する複数言語モデル選択手順と、選択された複数の言語モデルを入力して混合言語モデルを生成する混合言語モデル生成手順と、音響モデルと、生成された言語モデルを用いて、抽出された音声特徴量に対して照合を行い音声認識結果を出力する照合手順とを実現させることにより、学習用テキストデータが少量であることによって生じる言語モデルのゼロ頻度問題やスパースネスの問題を軽減でき、木構造クラスタ別言語モデルから複数選択した木構造クラスタ言語モデルによって混合言語モデルを生成して、音声認識に用いるので、さらに認識精度が高い音声認識ができると共に、認識対象の1発声が複数の分野や場面・状況を含む場合であっても、複数の分野や場面・状況の言語特徴を学習した言語モデルを選択し混合言語モデルを生成して音声認識に用いるので、認識性能が高い音声認識ができる効果があ

る。

【0115】この発明によれば、音声認識プログラムの複数言語モデル選択手順が、木構造クラスタ別言語モデルにおける最も下層の葉ノードのクラスタ別言語モデルから複数の言語モデルを選択することにより、言語モデルを選択する際の演算量を削減できるという効果が得られる。

【図面の簡単な説明】

【図1】 この発明の実施の形態1による言語モデル生成装置の構成を示すブロック図である。

【図2】 この発明の実施の形態1による言語モデル生成装置における言語モデル生成方法を示すフローチャートである。

【図3】 この発明の実施の形態1による学習用テキストデータ木構造クラスタリングの説明図である。

【図4】 この発明の実施の形態1による木構造クラスタ別の言語モデル生成の説明図である。

【図5】 この発明の実施の形態2による言語モデル生成装置の構成を示すブロック図である。

【図6】 この発明の実施の形態2による言語モデル生成装置における言語モデル生成方法を示すフローチャートである。

【図7】 この発明の実施の形態3による音声認識装置の構成を示すブロック図である。

【図8】 この発明の実施の形態3による音声認識装置における音声認識方法を示すフローチャートである。

【図9】 この発明の実施の形態4による音声認識装置の構成を示すブロック図である。

【図10】 この発明の実施の形態4による音声認識装置における音声認識方法を示すフローチャートである。

【図11】 この発明の実施の形態5による音声認識装\*

\* 置の構成を示すブロック図である。

【図12】 この発明の実施の形態5による音声認識装置における音声認識方法を示すフローチャートである。

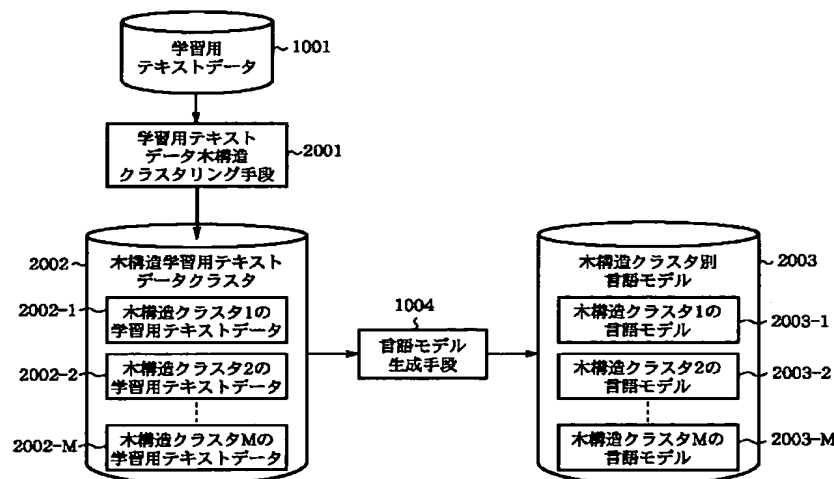
【図13】 従来の言語モデル生成装置の構成を示すブロック図である。

【図14】 従来の音声認識装置の構成を示すブロック図である。

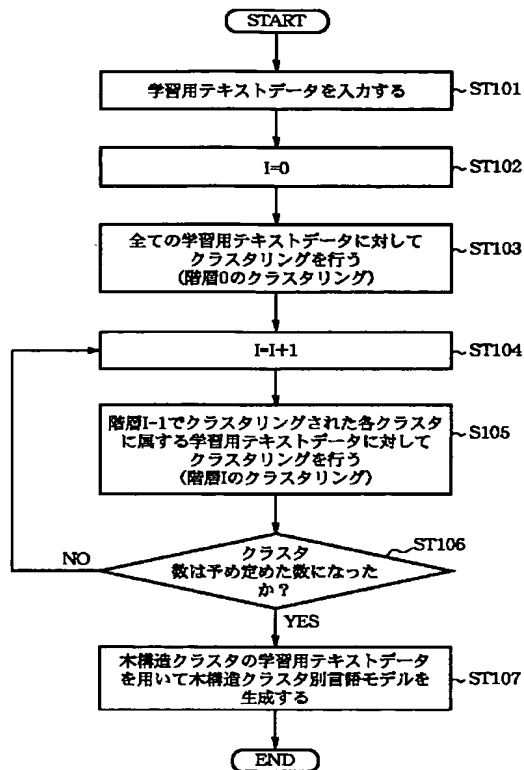
【符号の説明】

1001 学習用テキストデータ、1004 言語モデル生成手段、1101 認識対象音声、1102 音声特徴量抽出手段、1103 音響モデル、1104 言語モデル選択手段、1105 照合手段、1106 音声認識結果、2001 学習用テキストデータ木構造クラスタリング手段、2002 木構造学習用テキストデータクラスタ、2002-1 木構造クラスタ1の学習用テキストデータ、2002-2 木構造クラスタ2の学習用テキストデータ、2002-M 木構造クラスタMの学習用テキストデータ、2003 木構造クラスタ別言語モデル、2003-1 木構造クラスタ1の言語モデル、2003-2 木構造クラスタ2の言語モデル、2003-M 木構造クラスタMの言語モデル、3001 言語モデル補間手段、3002 補間処理された木構造クラスタ別言語モデル、3002-1 補間処理された木構造クラスタ1の言語モデル、3002-2 補間処理された木構造クラスタ2の言語モデル、3002-M 補間処理された木構造クラスタMの言語モデル、5001 複数言語モデル選択手段、5002 混合言語モデル生成手段、6001 葉ノードのクラスタ別言語モデル、6001-1 葉ノードクラスタ1の言語モデル、6001-2 葉ノードクラスタ2の言語モデル、6001-L 葉ノードクラスタLの言語モデル。

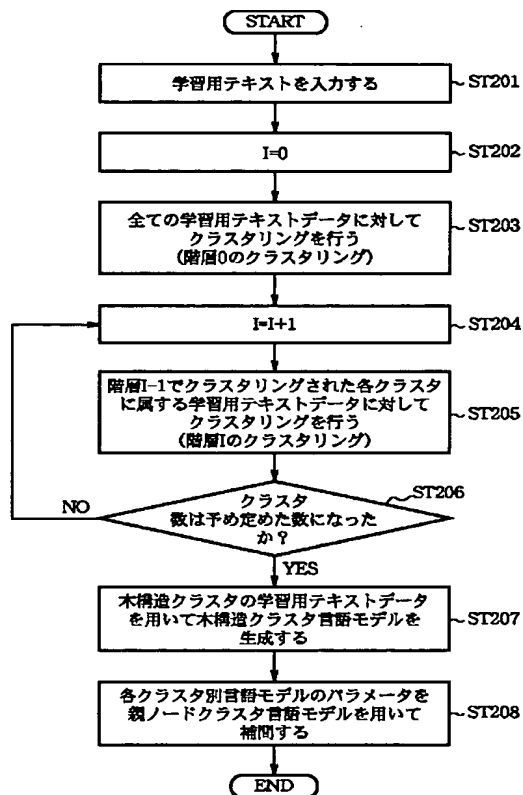
【図1】



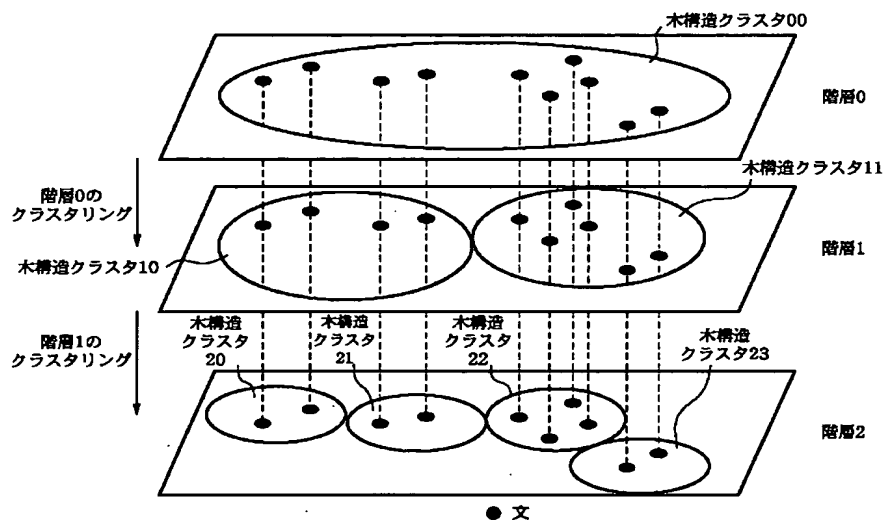
【図2】



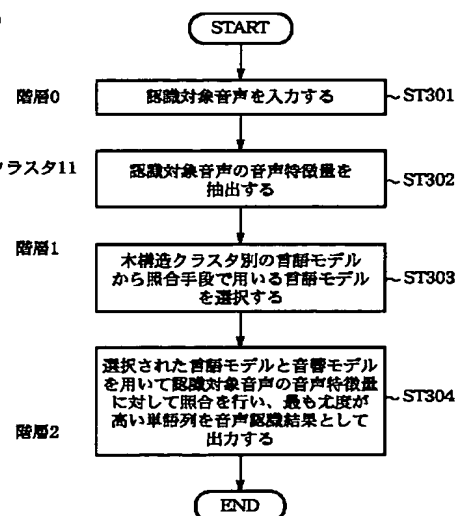
【図6】



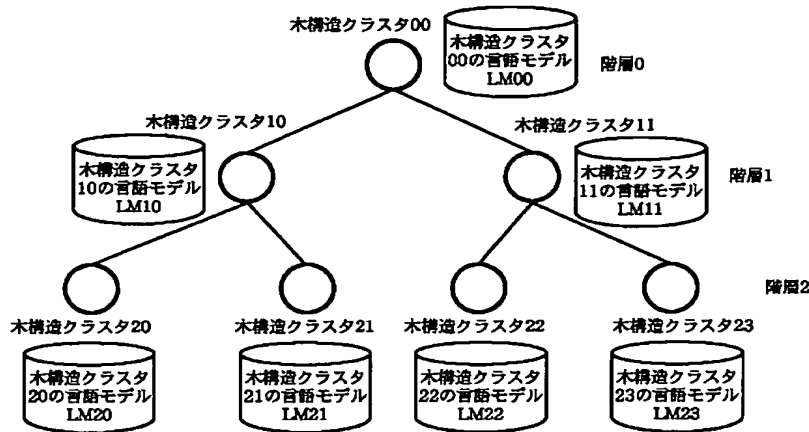
【図3】



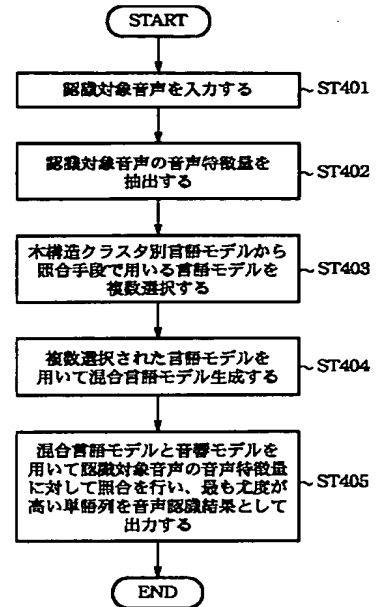
【図8】



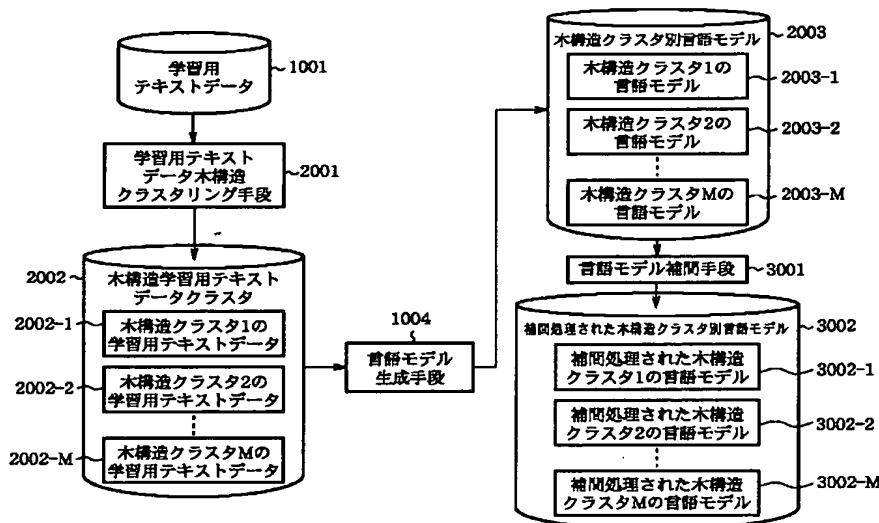
【図4】



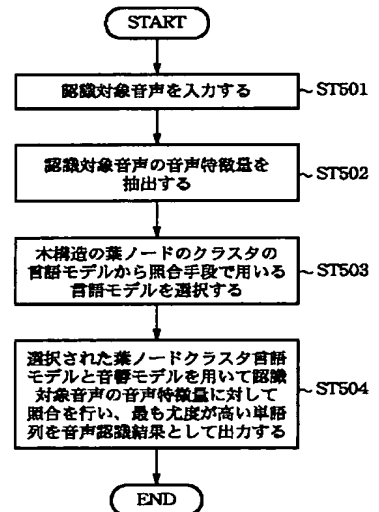
【図10】



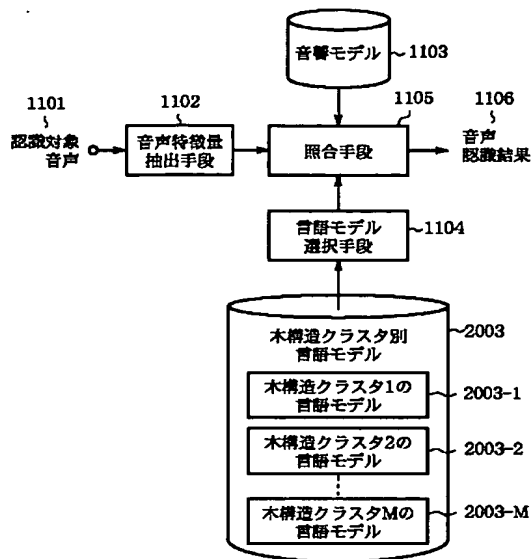
【図5】



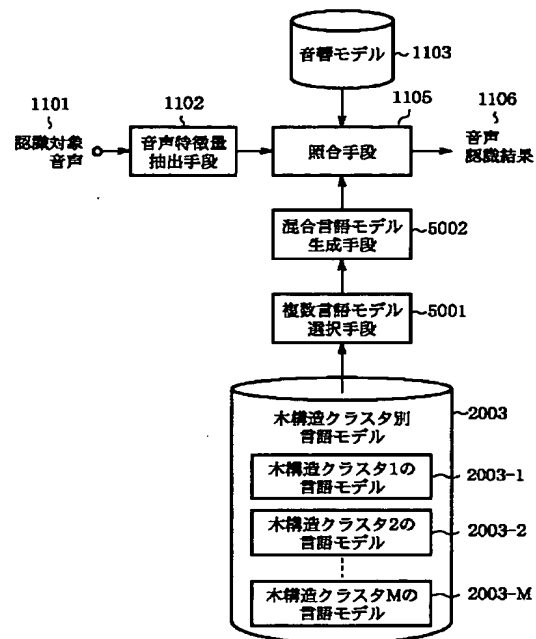
【図12】



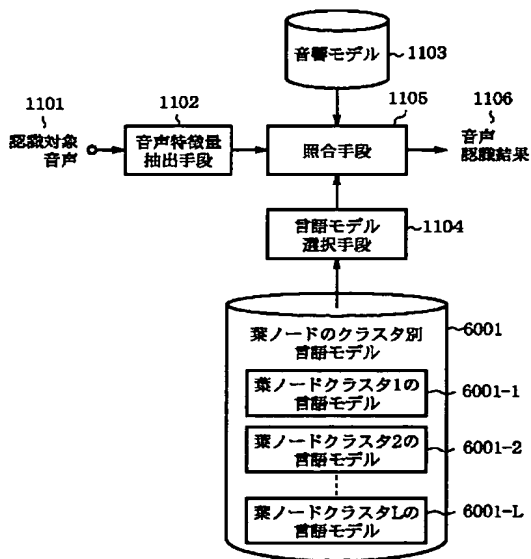
【図7】



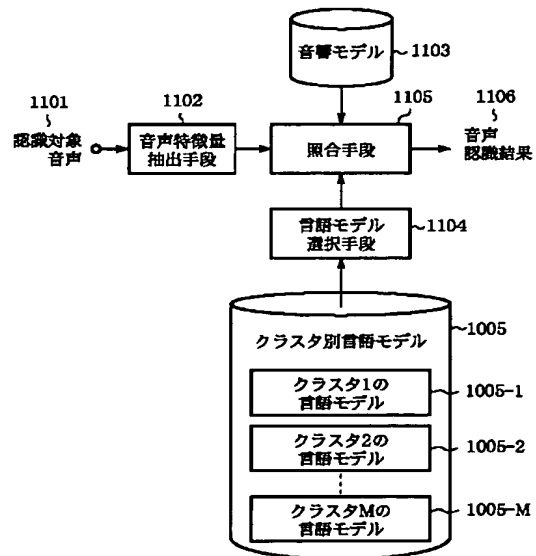
【図9】



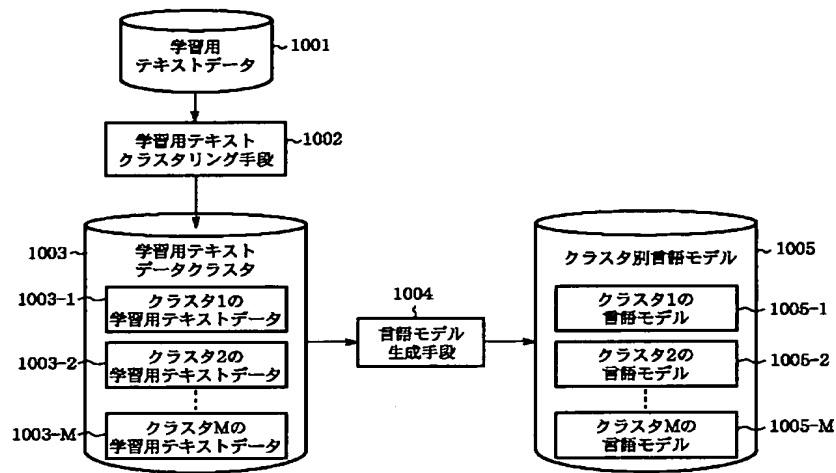
【図11】



【図14】



【図13】



フロントページの続き

- (54)【発明の名称】 言語モデル生成装置及びこれを用いた音声認識装置、言語モデル生成方法及びこれを用いた音声認識方法、並びに言語モデル生成プログラムを記録したコンピュータ読み取り可能な記録媒体及び音声認識プログラムを記録したコンピュータ読み取り可能な記録媒体



*This Page Blank (uspto)*



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**This Page Blank (uspto)**